

Generating correlated random vector by polynomial normal transformation

Qing Xiao

Received: date / Accepted: date

Abstract This paper develops a polynomial normal transformation model, whereby various non-normal probability distributions can be simulated by the standard normal distribution. Two methods are presented to determine the coefficients of polynomial model: (1) probability weighted moment (PWM) matching (2) percentile matching. Compared to the existing raw moment or L-moment matching, the proposed methods are more computationally convenient, and can be used to estimate the coefficients of polynomial model with a higher degree. Furthermore, for two correlated random variables, a polynomial equation is derived to estimate the equivalent correlation coefficient in standard normal space, and random vector with non-normal marginal distributions and prescribed correlation matrix can be generated. Finally, numerical examples are worked to demonstrate the proposed method.

Keywords Polynomial normal transformation · NORTA algorithm · Probability weighted moment · Percentile · Correlation coefficient

1 Introduction

In order to generate non-normal random variables, many families of distributions have been proposed. [Tadikamalla(1980)] summarizes the related works, and compares six procedures in terms of the efficiency, simplicity and generality. Among these approaches, two standard normal distribution based transformation models: Johnson system and third-order polynomial normal transformation (TPNT) technique, demonstrate the potential utility for generating correlated random variables.

[Johnson(1949)] discusses the feasibility of simulating continuous non-normal distributions by standard normal distribution, and develops a four-parameter transformation system. The Johnson system comprises four transformation models: S_N (Normal distribution), S_L (Lognormal family), S_U (Unbounded family) and S_B (Bounded family), which can be employed to approximate many non-normal distributions. For the detailed introduction to Johnson system, one can refer to literature[DeBroy et al(1989)].

Qing Xiao
School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200072, China
E-mail: xaoshaoying@shu.edu.cn

Johnson system has been widely used to generate non-normal random numbers, but it shows inadequacy of accommodating the dependence among random variables. If samples of the correlated random vector are available, the Johnson system is workable. By normalizing each sample value through the corresponding transformation model, the correlation matrix in the normal space can be calculated directly (see Section 3.2 of reference[Stanfield et al(1996)]). However, this approach cannot be used to generate random vector with assigned correlation matrix. To achieve this goal, it is necessary to analytically estimate the correlation coefficient ρ_z in normal space for a given correlation coefficient ρ_x between two random variables.

Compared with Johnson system, the dependence among random variables can be easily handled by TPNT technique. For a given ρ_x , the analytical equation can be derived to estimate ρ_z , and random vector with desired correlation matrix can be conveniently generated[Vale and Maurelli(1983), Headrick and Sanwilowsky(1999)]. The TPNT is firstly introduced by [Fleishman(1978)], this technique generates non-normal deviates by a polynomial transformation of standard normal ones. A detailed investigation of TPNT is given in reference[Chen and Tung(2003)], it is found there are many probability distributions that TPNT technique cannot simulate.

To expand the generality of the polynomial transformation technique, a fifth-order polynomial normal transformation (FPNT) technique is developed in reference [Headrick(2002)]. Due to the ability to control the extra fifth and sixth order moments, FPNT can approximate some distributions that are difficult for TPNT[Headrick and Kowalchuk(2007)], but it requires tedious mathematical computation to determine the coefficients of FPNT. By equating the first six raw moments of FPNT model with those of the modeled distribution, a system of nonlinear equations is established. The coefficients are estimated by solving the equations numerically. In pursuit of a simpler parameter estimation method, L-moments are employed to characterize FPNT model[Headrick(2011)], and the analytical expression of coefficients are obtained. However, because of the tedious mathematical derivation, this L-moments based method is unable to estimate the coefficients of a polynomial model with a higher degree, which may be used to simulate more non-normal distributions.

In this paper, the polynomial normal transformation model is extended to a much higher degree. Two methods: probability weighted moment (PWM) matching and percentile matching, are developed to assess the coefficients of polynomial model. By equating PWMs of the polynomial model with those of the modeled distribution, a system of linear equations solved for the coefficients is established. For the percentile matching, the polynomial model can be established by a least square method, and a numerical investigation is also performed to check the generality and accuracy of this method.

Furthermore, using the product moments formulae of two correlated standard normal variables [Pearson and Young(1918)], the polynomial equation of ρ_z is derived. For a given ρ_x between two correlated random variables, an appropriate value of ρ_z is conveniently determined, and random vector with assigned correlation matrix can be generated from standard normal deviates. Through numerical examples, it makes evident that non-normal distributions can be well simulated by the polynomial model, and random vector with prescribed correlation structure can be generated by the proposed procedures.

2 Polynomial normal transformation model

Let X be a continuous random variable with cumulative distribution function (CDF) $F(X)$, let Z be a standard normal random variable with CDF $\Phi(Z)$. Equate the two CDFs:

$$F(X) = \Phi(Z) \quad (1)$$

Then, X is expressed as:

$$X = F^{-1}[\Phi(Z)] \quad (2)$$

where $F^{-1}(\cdot)$ denotes the inverse CDF of X .

Approximate the function $F^{-1}[\Phi(Z)]$ by a polynomial of degree n :

$$X \simeq \sum_{k=0}^n a_k Z^k \quad (3)$$

where a_k ($k = 0, \dots, n$) are the undetermined coefficients. Setting $n = 3$ or 5 would yield the TPNT or FPNT model.

Since both $F^{-1}(\cdot)$ and $\Phi(Z)$ are monotonic increasing functions, so is the composite function $F^{-1}[\Phi(Z)]$. According to Weierstrass approximation theorem[Saxe(2002)], the monotonic smooth function $F^{-1}[\Phi(Z)]$ can be well approximated by a polynomial of Z over a closed interval. While, the major problem is to determine a set of coefficients a_k ($k = 0, \dots, n$), such that the random variable X would be well simulated by the standard normal variable Z . Most often, the researchers resort to the raw moment matching method[Fleishman(1978),Headick(2002)], which requires tedious mathematical computation. For a polynomial of degree 5, the resulted system of nonlinear equations is very complicated (see Appendix A of reference[Headick(2002)]), let alone for a polynomial model of higher degree. In this paper, the PWM matching and percentile matching are introduced to simplify the coefficients estimation problem, and the polynomial model is extended to higher degree, allowing for the simulation of more probability distributions.

2.1 Probability weighted moment matching method

The PWM of X is defined as [Greenwood et al(1979)]:

$$M_{r,p,s} = E\{[F(X)]^r \cdot X^p \cdot [1 - F(X)]^s\} \quad (4)$$

For computational convenience, a particular type of PWM, $\beta_r = M_{r,1,0}$, is considered:

$$\beta_r = E\{[F(X)]^r \cdot X\} = \int_{-\infty}^{+\infty} F^r(X) \cdot X \cdot f(X) dX \quad (5)$$

where $f(X)$ is the probability density function (PDF).

If the PDF of X is unknown, only a set of data is available. Sort the sample into ascending order: $x_1 \leq \dots \leq x_i \leq \dots \leq x_m$, the unbiased estimate of β_r is [Hosking et al(1985)Hosking, Wallis, and Wood]:

$$\beta_r = \frac{1}{m} \sum_{i=r+1}^m \frac{(i-1)(i-2) \cdots (i-r)}{(m-1)(m-2) \cdots (m-r)} x_i \quad (6)$$

Substitute Eq.(3) into Eq.(5), the PWM of polynomial model is:

$$\begin{aligned}
\beta_r &= \int_{-\infty}^{+\infty} \Phi^r(Z) \cdot \left(\sum_{k=0}^n a_k Z^k \right) \cdot \varphi(Z) dZ \\
&= \sum_{k=0}^n a_k \int_{-\infty}^{+\infty} \Phi^r(Z) \cdot Z^k \cdot \varphi(Z) dZ \\
&= \sum_{k=0}^n a_k M_{r,k,0}^z \\
M_{r,k,0}^z &= \int_{-\infty}^{+\infty} \Phi^r(Z) \cdot Z^k \cdot \varphi(Z) dZ
\end{aligned} \tag{7}$$

where $\varphi(Z)$ is the PDF. Since $\Phi(Z)$ and $\varphi(Z)$ are known functions, $M_{r,k,0}^z$ can be integrated numerically.

For an n th order polynomial model, perform the first $(n+1)$ PWMs matching, the following system of linear equations is established:

$$\begin{pmatrix} M_{0,0,0}^z & \cdots & M_{0,k,0}^z & \cdots & M_{0,n,0}^z \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ M_{r,0,0}^z & \cdots & M_{r,k,0}^z & \cdots & M_{r,n,0}^z \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ M_{n,0,0}^z & \cdots & M_{n,k,0}^z & \cdots & M_{n,n,0}^z \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ \vdots \\ a_k \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_r \\ \vdots \\ \beta_n \end{pmatrix} \tag{8}$$

Calculate the first $(n+1)$ PWMs of X :

$$\beta_r = \int_{-\infty}^{+\infty} F^r(X) \cdot X \cdot f(X) dX = \int_0^1 F^{-1}(p) \cdot p^r dp \tag{9}$$

Substitute these PWMs into Eq.(8), solving the system of linear equations gives the coefficients a_k .

Intuitively, a polynomial of higher degree allows for the control of higher order moments, and should give a better approximation of the target distribution. However, as the degree of the polynomial n increases, the coefficient matrix becomes near singular. Below is the determinant of the coefficient matrix:

Table 1: The determinant of coefficient matrix

n	Determinant	n	Determinant
1	0.2821	7	3.3321×10^{-14}
2	0.0259	8	7.1266×10^{-18}
3	8.2291×10^{-4}	9	5.7686×10^{-22}
4	9.3220×10^{-6}	10	1.7762×10^{-26}
5	3.8458×10^{-8}	11	2.0880×10^{-31}
6	5.8600×10^{-11}	12	9.4023×10^{-37}

For a system of linear equations with a near singular coefficient matrix, the solution is very sensitive

to small errors in the estimation of $M_{r,k,0}^z$ or β_r . Thus, when modeling theoretical distributions, a polynomial model of degree higher than 12 is not recommended, otherwise, the near singular coefficient matrix would yield biased values of a_k , giving rise to an undesirable simulation. Similarly, when used for data fitting, a polynomial of lower degree is more preferable, in case of the error caused by the high-order PWMs drawn from the samples (In [Zou and Xiao(2014)], a 9th-order polynomial model is employed).

2.2 Percentile matching method

For a polynomial model, the following requirement should be satisfied:

$$x_p = a_0 + a_1 z_p + \cdots + a_n z_p^n \quad (10)$$

where p is a percentage ($p \in [0, 1]$), x_p and z_p are the corresponding percentiles from the target distribution $F(X)$ and the standard normal distribution $\Phi(Z)$:

$$x_p = F^{-1}(p) \quad z_p = \Phi^{-1}(p) \quad (11)$$

Select m values of p_i , $m \geq (n + 1)$, calculate the associated percentiles: z_{p_i} and x_{p_i} ($i = 1, \dots, m$). Based on m pair values of (z_{p_i}, x_{p_i}) , the polynomial model can be established by a least square method.

Here, a heuristic method is put forward to establish the polynomial by percentile matching:

1. Choose 14, 16 and 15 percentage points of p_i evenly over the interval $[\alpha, 0.01)$, $[0.01, 0.99)$ and $[0.99, 1 - \alpha]$ respectively, (α is a small value of percentage).
2. For each percentage p_i , calculate the corresponding percentiles of $F(x)$ and $\Phi(z)$ as stated in Eq.(11).
3. Using the resulted 45 pair values of (z_{p_i}, x_{p_i}) , a 19th order polynomial can be established by a least square method.

To check the performance of the proposed method, the absolute relative error between the percentile from the original distribution and the one from polynomial model is calculated:

$$\varepsilon_{p_i} = \left| \frac{x_{p_i}^* - x_{p_i}}{x_{p_i}} \right| \times 100[\%], \quad x_{p_i} = F^{-1}(p_i), \quad (12)$$

$$x_{p_i}^* = \sum_{k=0}^n a_k z_{p_i}^k, \quad z_{p_i} = \Phi^{-1}(p_i).$$

Testing for various probability distributions in Matlab, such as: Gamma distribution $\text{Gamma}(a, b)$, Beta distribution $\text{Beta}(a, b)$, F-distribution $F(d_1, d_2)$, Lognormal distribution $\ln N(a, b)$, Weibull distribution $\text{Weibull}(a, b)$, Uniform distribution $U(a, b)$, Gumbel distribution $\text{Gumble}(a, b)$, Logistic distribution $\text{Logistic}(a, b)$, T-distribution $T(\nu)$, Chi squared distribution $\text{chi2}(k)$, Rayleigh distribution $\text{Rayleigh}(\sigma)$ and Exponential distribution $\text{Exp}(\lambda)$. the results are shown in Table 2.

As shown in Table 2, the tested non-normal distributions are well approximated by the 19th order polynomial model. Taking the Gamma distribution $\text{Gamma}(a, b)$ for example, the error ε_{p_i} for the family of Gamma distribution with parameters $1 \leq a \leq 100$, $1 \leq b \leq 100$ is all less than 0.92%, demonstrating a satisfactory accuracy.

However, if the polynomial model is employed for data fitting, a large number of samples should be provided to estimate a precise value of the percentile x_{p_k} . With a small to moderate sample size, the PWM matching method is more favourable.

Table 2: The absolute relative error of 19th order polynomial model for simulating non-normal distributions

Probability distribution	Parameters	Probit range $[\alpha, 1 - \alpha]$	Maximum value of $\varepsilon_{p_k}(\%)$
Gamma	$1 \leq a \leq 100$ $1 \leq b \leq 100$	$[10^{-4}, 1 - 10^{-4}]$	0.92
Beta	$2 \leq a \leq 20$ $2 \leq b \leq 20$	$[10^{-4}, 1 - 10^{-4}]$	2.3×10^{-4}
Beta	$1 \leq a \leq 2$ $1 \leq b \leq 2$	$[10^{-3}, 1 - 10^{-3}]$	6×10^{-6}
Weibull	$1 \leq a \leq 100$ $1 \leq b \leq 100$	$[10^{-4}, 1 - 10^{-4}]$	0.92
Lognormal	$0.01 \leq a \leq 200$ $0.01 \leq b \leq 4$	$[10^{-4}, 1 - 10^{-4}]$	0.060
F	$4 \leq d_1 \leq 100$ $4 \leq d_2 \leq 100$	$[10^{-4}, 1 - 10^{-4}]$	0.090
Uniform	$a = 0 \ b = 1$	$[10^{-3}, 1 - 10^{-3}]$	0.0035
Gumbel	$a = 0 \ b = 1$	$[10^{-4}, 1 - 10^{-4}]$	0.013
Logistic	$a = 0 \ b = 1$	$[10^{-4}, 1 - 10^{-4}]$	6.6×10^{-5}
T	$1 \leq \nu \leq 100$	$[10^{-4}, 1 - 10^{-4}]$	0.064
Chi squared	$2 \leq k \leq 100$	$[10^{-4}, 1 - 10^{-4}]$	0.94
Rayleigh	$\sigma = 1$	$[10^{-4}, 1 - 10^{-4}]$	0.0063
Exponential	$\lambda = 1$	$[10^{-4}, 1 - 10^{-4}]$	0.95

3 Evaluating the correlation coefficient

Suppose X_1 and X_2 are two correlated random variables. Let X_1, X_2 both be approximated by a polynomial of degree n :

$$\begin{aligned} X_1 &= a_{1,0} + a_{1,1}Z_1 + \cdots + a_{1,n}Z_1^n \\ X_2 &= a_{2,0} + a_{2,1}Z_2 + \cdots + a_{2,n}Z_2^n \end{aligned} \quad (13)$$

To ensure a desired correlation coefficient ρ_x between X_1 and X_2 , a suitable correlation coefficient ρ_z between Z_1 and Z_2 should be determined. This problem is known as the Nataf transformation [Liu and Der Kiureghian, Lebrun and Dutfoy(2009)] or Normal-To-Anything (NORTA) algorithm [Cario and Nelson(1997), Chen(2001)], which requires solving the following integral equation:

$$\rho_x \sigma_1 \sigma_2 + \mu_1 \mu_2 = E[X_1 X_2] = \int \int F_1^{-1}[\Phi(Z_1)] F_2^{-1}[\Phi(Z_2)] \phi(Z_1, Z_2, \rho_z) dZ_1 dZ_2, \quad (14)$$

where μ_i, σ_i are the mean and standard deviation of X_i ($i = 1, 2$), $\phi(Z_1, Z_2, \rho_z)$ is the joint PDF of two normal random variables with correlation coefficient ρ_z . Generally, this equation is difficult to be analytically solved. While, using the polynomial normal transformation technique, the problem can be greatly simplified.

The formulae of the product moments of two correlated standard normal variables are as follows [Pearson and Young(1918)]:

$$\begin{aligned}
E(Z_1^{2s} Z_2^{2t}) &= \frac{(2s)!(2t)!}{2^{s+t}} \sum_{j=0}^{\min(s,t)} \frac{(2\rho_z)^{2j}}{(s-j)!(t-j)!(2j)!} \\
E(Z_1^{2s+1} Z_2^{2t+1}) &= \rho_z \frac{(2s+1)!(2t+1)!}{2^{s+t}} \cdot \sum_{j=0}^{\min(s,t)} \frac{(2\rho_z)^{2j}}{(s-j)!(t-j)!(2j+1)!} \\
E(Z_1^{2s} Z_2^{2t+1}) &= E(Z_1^{2s+1} Z_2^{2t}) = 0 \\
E(Z_i^{2s}) &= \frac{(2s)!}{2^s \cdot s!} \quad i = 1, 2.
\end{aligned} \tag{15}$$

where s and t are nonnegative integers.

Then, the following derivation is performed:

$$\begin{aligned}
\rho_x \sigma_1 \sigma_2 + \mu_1 \mu_2 &= E(X_1 X_2) = E \left[\left(\sum_{i=0}^n a_{1,i} Z_1^i \right) \left(\sum_{j=0}^n a_{2,j} Z_2^j \right) \right] \\
&= \begin{pmatrix} a_{1,0} & a_{1,1} & \cdots & a_{1,n} \end{pmatrix} E \begin{pmatrix} 1 & Z_2 & \cdots & Z_2^n \\ Z_1 & Z_1 Z_2 & \cdots & Z_1 Z_2^n \\ \vdots & \vdots & \cdots & \vdots \\ Z_1^n & Z_1^n Z_2 & \cdots & Z_1^n Z_2^n \end{pmatrix} \begin{pmatrix} a_{2,0} \\ a_{2,1} \\ \vdots \\ a_{2,n} \end{pmatrix}.
\end{aligned} \tag{16}$$

Substitute Eq.(15) into Eq.(16), and perform matrix multiplication, a polynomial equation in ρ_z of degree n can be obtained:

$$\rho_x \sigma_1 \sigma_2 + \mu_1 \mu_2 = b_n \rho_z^n + \cdots + b_i \rho_z^i + \cdots + b_1 \rho_z + b_0, \tag{17}$$

where b_i ($i = 0, \dots, n$) are sums of product of a_{1,k_1} and a_{2,k_2} . A 19th-order polynomial is presented in Appendix.

Then, the correlation coefficient ρ_x is expressed as a polynomial function of ρ_z . As indicated by Lemma 3 in reference[Liu and Der Kiureghian(1986)], the feasible correlation coefficient ρ_x cannot take any value between -1 and 1 , and is located in a subinterval of $[-1, 1]$, i.e. $-1 \leq \underline{\rho_x} \leq \rho_x \leq \overline{\rho_x} \leq 1$. Based on the derived polynomial equation, the lower and upper bounds of ρ_x can be determined by setting $\rho_z = -1, 1$.

Furthermore, some properties of the function relationship between ρ_z and ρ_x have been proved:

1. ρ_x is a strictly increasing function of ρ_z (Lemma 1 in reference[Liu and Der Kiureghian(1986)]).
2. $\rho_z \geq 0$ (≤ 0) implies $\rho_x \geq 0$ (≤ 0) (Proposition 1 in reference [Cario and Nelson(1997)]), that's to say, $\rho_z \rho_x \geq 0$.

Thus, for a given correlation coefficient ρ_x between X_1 and X_2 , the associated ρ_z can be estimated by solving the polynomial equation, and a valid solution is restricted by the following conditions:

$$-1 \leq \rho_z \leq 1 \quad \text{and} \quad \rho_z \rho_x \geq 0 \tag{18}$$

As long as X_1 and X_2 can be well approximated by the polynomial model, ρ_z for a given ρ_x can be estimated with a satisfactory accuracy.

Suppose $\mathbf{X} = (X_1, \dots, X_i, \dots, X_m)^T$ is an m -dimensional random vector with correlation matrix $\mathbf{R}_\mathbf{X}$, the steps for generating \mathbf{X} are as follows:

- (1) Approximate each random variable X_i by the polynomial model, determine the coefficients by the PWM matching method or percentile matching method.
- (2) Calculate $\rho_z(i, j)$ ($i \neq j$) for each entry $\rho_x(i, j)$ in \mathbf{R}_X , obtain \mathbf{R}_Z , which is the equivalent correlation matrix in the standard normal space.
- (3) Generate m -dimensional independent standard normal vector: $\mathbf{U} = (U_1, \dots, U_i, \dots, U_m)^T$, transform \mathbf{U} into the correlated standard normal vector $\mathbf{Z} = (Z_1 \dots Z_i \dots Z_m)^T$ with correlation matrix \mathbf{R}_Z by performing $\mathbf{Z} = \mathbf{L}\mathbf{U}$. \mathbf{L} is the lower triangular matrix resulted from Cholesky decomposition, that is, $\mathbf{R}_Z = \mathbf{L}\mathbf{L}^T$.
- (4) Convert Z_i into X_i through the marginal transformation $X_i = F_i^{-1}[\Phi(Z_i)]$, the random vector \mathbf{X} with the prescribed marginal distributions and correlation matrix \mathbf{R}_X is obtained.

The procedures can be expressed as follows:

$$\begin{pmatrix} U_1 \\ \vdots \\ U_i \\ \vdots \\ U_m \end{pmatrix} \xrightarrow[\substack{\mathbf{R}_Z = \mathbf{L}\mathbf{L}^T \\ \downarrow \\ \mathbf{Z} = \mathbf{L}\mathbf{U}}]{\quad} \begin{pmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_m \end{pmatrix} \xrightarrow{X_i = F_i^{-1}(Z_i)} \begin{pmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_m \end{pmatrix} \quad (19)$$

As long as the marginal distribution $F_i(X_i)$ can be well approximate by polynomial model, and the correlation matrix \mathbf{R}_X is a positive definite symmetric matrix (see Proposition 3 in reference[Cario and Nelson(1997)]), the correlated random vector \mathbf{X} can be generated by the proposed method.

4 Numerical example

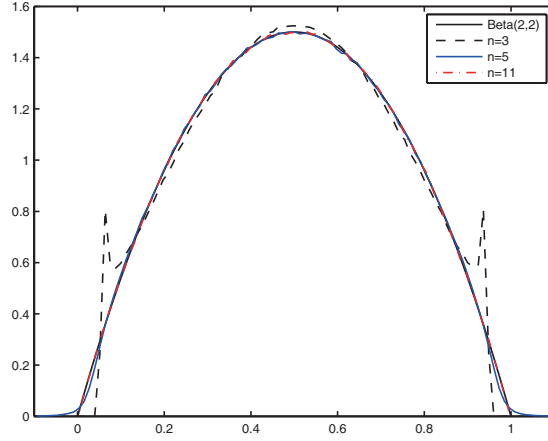
4.1 Simulating non-normal distributions

The polynomial models of degree 3, 5, 11 are employed to simulate the Beta distribution $Beta(2, 2)$ respectively. The coefficients are determined by PWM matching method in Section 2.1, and the coefficients of the 11th polynomial model are shown in Table 3. The PDFs of the random numbers generated by polynomial models are depicted in Fig. 1.

Table 3: The coefficients of 11th order polynomial model— $Beta(2, 2)$

k	a_k	k	a_k
11	$-2.76217093165881 \times 10^{-8}$	5	0.00120213182254751
10	$-8.53583161405554 \times 10^{-10}$	4	$1.41667915745929 \times 10^{-7}$
9	$1.75657290203781 \times 10^{-6}$	3	-0.0192416046002625
8	$1.63734319656302 \times 10^{-8}$	2	$-6.4707982562276 \times 10^{-8}$
7	-0.0000555670500494067	1	0.265961312977451
6	$-8.51019206906232 \times 10^{-8}$	0	0.500000003658777

10^4 values of p_i are evenly chosen in the probit range $[0.001, 0.999]$. Calculate ε_{p_i} for each percentage p_i ($i = 1, \dots, 10^4$) as stated in Eq.(12), the average value, the minimum value, and the maximum value of ε_{p_i} are presented in Table 4.

Fig. 1: The PDFs of Beta distribution $Beta(2, 2)$ Table 4: The values of ε_{p_i} for simulating $Beta(2, 2)$

Degree (n)	Average(%)	Minimum(%)	Maximum(%)
3	1.2	1.8×10^{-5}	389
5	0.15	8.0×10^{-6}	54
11	4.1×10^{-4}	2.0×10^{-9}	0.095

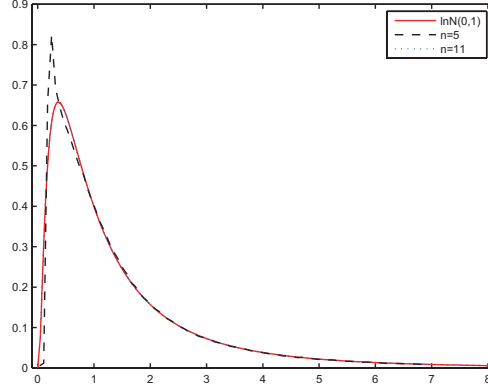
As shown in Fig. 1 and Table 4, due to the ability of controlling higher order moments, a polynomial model of higher degree provides a more accurate approximation, the PDF given by the 11th order polynomial model is in close proximity with Beta distribution.

The TPNT model cannot provide a good approximation for the Lognormal distribution $\ln N(0, 1)$. Here, the polynomial models of degree 5 and 11 are employed to simulate $\ln N(0, 1)$. 17 percentage points p_i are evenly chosen in the interval $[0.001, 0.999]$. Estimate z_{p_i} and x_{p_i} as stated in Eq.(11). With the resulted 17 pair values of (z_{p_i}, x_{p_i}) , the coefficients of the polynomial model are determined by the least square method. The coefficients of the 11th polynomial model are shown in Table 5.

Table 5: The coefficients of 11th order polynomial model— $-\ln N(0, 1)$

k	a_k	k	a_k
11	$2.70587705587477 \times 10^{-8}$	5	0.008333335056026873
10	$3.07095310533251 \times 10^{-7}$	4	0.0416665771671574
9	$2.75245190929274 \times 10^{-6}$	3	0.166666657138832
8	$2.46853662677270 \times 10^{-5}$	2	0.500000016464362
7	0.000198405765895619	1	1.00000000118744
6	0.00138905069923553	0	0.999999999545197

Fig. 2 shows the PDFs of the Lognormal distribution and its polynomial model analogs. Calculate ε_{p_i} as in the former case, the results are summarized in Table 6. As can be seen, the 11th polynomial model gives better performance than the 5th one.

Fig. 2: The PDFs of Lognormal distribution $\ln N(0, 1)$ Table 6: The values of ε_p for simulating $\ln N(0, 1)$

Degree (n)	Average(%)	Minimum(%)	Maximum(%)
5	3.6	7.2×10^{-5}	167
11	4.6×10^{-4}	0	0.087

4.2 Generating correlated random vector

Suppose X_1 and X_2 both follow the Lognormal distribution $\ln N(0, 1)$. In this case, the function relationship of ρ_x and ρ_z is as follows [Mostafa and Mahmoud(1964)]:

$$\rho_x = \frac{\exp(\rho_z) - 1}{e - 1} \quad (20)$$

For a given ρ_x , ρ_z resulted from Eq.(20) is regarded as benchmark.

The foregoing 11th order polynomial model is used to simulate X_1 and X_2 , of which the coefficients are presented in Table 5. In Appendix, the coefficients a_k ($k = 12, \dots, 19$) are set to be 0. The results are summarized in Table 7. The function curves are shown in Fig. 3.

Table 7: The correlation coefficients ρ_z between $\ln N(0, 1)$ and $\ln N(0, 1)$

ρ_x	ρ_z (Exact value)	ρ_z (Equation)
-0.3	-0.725	-0.725
-0.1	-0.189	-0.189
0.1	0.159	0.159
0.3	0.416	0.416
0.5	0.620	0.620
0.7	0.790	0.790
0.9	0.935	0.935

From Table 7 and Fig. 3, it can be seen that the proposed method yields results with satisfactory accuracy.

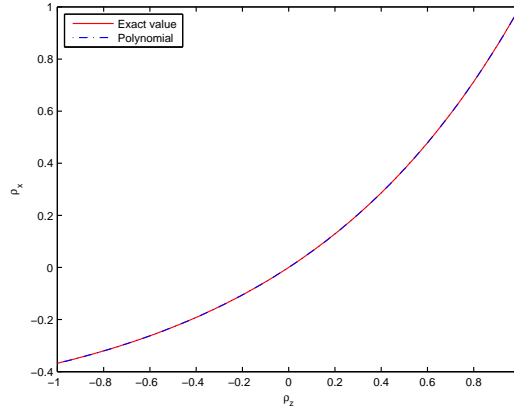


Fig. 3: The function relationship between ρ_z and ρ_x

Finally, an example of generating correlated random vector is worked. Suppose $\mathbf{X} = (X_1, X_2, X_3)^T$ is a three dimensional random vector, X_1, X_2, X_3 follow the standard normal distribution $N(0, 1)$, Beta distribution $Beta(2, 2)$ and Lognormal distribution $lnN(0, 1)$, respectively. The desired correlation matrix \mathbf{R}_X is:

$$\mathbf{R}_X = \begin{pmatrix} 1 & 0.9 & 0.5 \\ 0.9 & 1 & 0.3 \\ 0.5 & 0.3 & 1 \end{pmatrix}$$

For the normal distribution with mean μ and standard deviation σ , the coefficients of the polynomial can be regarded as: $a_0 = \mu$, $a_1 = \sigma$ and $a_i = 0$ ($i = 2, \dots, n$). The Beta distribution and Lognormal distribution are both simulated by the 11th order polynomial model, the coefficients are tabulated in Table 3 and Table 5 respectively. The equivalent correlation matrix \mathbf{R}_Z in the normal space is:

$$\mathbf{R}_Z = \begin{pmatrix} 1 & 0.907 & 0.655 \\ 0.907 & 1 & 0.400 \\ 0.655 & 0.400 & 1 \end{pmatrix}$$

10^6 three-dimensional random vectors are generated as stated in Eq.(19). The correlation matrix of samples is:

$$\mathbf{R}_X^* = \begin{pmatrix} 1 & 0.900 & 0.499 \\ 0.900 & 1 & 0.300 \\ 0.499 & 0.300 & 1 \end{pmatrix}$$

Comparing \mathbf{R}_X^* with \mathbf{R}_X , it makes evident that the correlation matrix of samples is in close proximity to the desired one. Multivariate random variables with prescribed marginal distributions and correlation matrix are generated.

5 Conclusions

A polynomial normal transformation model is developed in this paper. The PWM matching method and the percentile matching method are introduced to determine the coefficients, whereby the

polynomial normal transformation model is extended to a higher degree. Comparing to the extant TPNT and FPNT models, the proposed model can preserve more statistical information of the target distribution and gives more accurate approximation.

For two correlated random variables, the polynomial equations for estimating ρ_z are also derived. It is shown that the original correlation coefficient ρ_x can be expressed as a polynomial function of ρ_z . For a given ρ_x , the associated ρ_z is estimated by solving the polynomial equation. The numerical examples demonstrate that non-normal distributions can be well approximated by the polynomial model, and random vector with prescribed correlation matrix can be generated by the proposed method.

Appendix

$$\begin{aligned}
b_{19} &= 121645100408832000a_{1,19}a_{2,19} \\
b_{18} &= 6402373705728000a_{1,18}a_{2,18} \\
b_{17} &= 355687428096000(a_{1,17} + 171a_{1,19})(a_{2,17} + 171a_{2,19}) \\
b_{16} &= 20922789888000(a_{1,16} + 153a_{1,18})(a_{2,16} + 153a_{2,18}) \\
b_{15} &= 1307674368000(a_{1,15} + 136a_{1,17} + 11628a_{1,19})(a_{2,15} + 136a_{2,17} + 11628a_{2,19}) \\
b_{14} &= 87178291200(a_{1,14} + 120a_{1,16} + 9180a_{1,18})(a_{2,14} + 120a_{2,16} + 9180a_{2,18}) \\
b_{13} &= 6227020800(a_{1,13} + 105a_{1,15} + 7140a_{1,17} + 406980a_{1,19}) \\
&\quad (a_{2,13} + 105a_{2,15} + 7140a_{2,17} + 406980a_{2,19}) \\
b_{12} &= 479001600(a_{1,12} + 91a_{1,14} + 5460a_{1,16} + 278460a_{1,18}) \\
&\quad (a_{2,12} + 91a_{2,14} + 5460a_{2,16} + 278460a_{2,18}) \\
b_{11} &= 39916800(a_{1,11} + 78a_{1,13} + 4095a_{1,15} + 185640a_{1,17} + 7936110a_{1,19}) \\
&\quad (a_{2,11} + 78a_{2,13} + 4095a_{2,15} + 185640a_{2,17} + 7936110a_{2,19}) \\
b_{10} &= 3628800(a_{1,10} + 66a_{1,12} + 3003a_{1,14} + 120120a_{1,16} + 4594590a_{1,18}) \\
&\quad (a_{2,10} + 66a_{2,12} + 3003a_{2,14} + 120120a_{2,16} + 4594590a_{2,18}) \\
b_9 &= 362880(a_{1,9} + 55a_{1,11} + 2145a_{1,13} + 75075a_{1,15} + 2552550a_{1,17} + 87297210a_{1,19}) \\
&\quad (a_{2,9} + 55a_{2,11} + 2145a_{2,13} + 75075a_{2,15} + 2552550a_{2,17} + 87297210a_{2,19}) \\
b_8 &= 40320(a_{1,8} + 45a_{1,10} + 1485a_{1,12} + 45045a_{1,14} + 1351350a_{1,16} + 41351310a_{1,18}) \\
&\quad (a_{2,8} + 45a_{2,10} + 1485a_{2,12} + 45045a_{2,14} + 1351350a_{2,16} + 41351310a_{2,18}) \\
b_7 &= 5040(a_{1,7} + 36a_{1,9} + 990a_{1,11} + 25740a_{1,13} + 675675a_{1,15} + 18378360a_{1,17} + 523783260a_{1,19}) \\
&\quad (a_{2,7} + 36a_{2,9} + 990a_{2,11} + 25740a_{2,13} + 675675a_{2,15} + 18378360a_{2,17} + 523783260a_{2,19}) \\
b_6 &= 720(a_{1,6} + 28a_{1,8} + 630a_{1,10} + 13860a_{1,12} + 315315a_{1,14} + 7567560a_{1,16} + 192972780a_{1,18}) \\
&\quad (a_{2,6} + 28a_{2,8} + 630a_{2,10} + 13860a_{2,12} + 315315a_{2,14} + 7567560a_{2,16} + 192972780a_{2,18}) \\
b_5 &= 120(a_{1,5} + 21a_{1,7} + 378a_{1,9} + 6930a_{1,11} + 135135a_{1,13} + 2837835a_{1,15} + 64324260a_{1,17} + 1571349780a_{1,19}) \\
&\quad (a_{2,5} + 21a_{2,7} + 378a_{2,9} + 6930a_{2,11} + 135135a_{2,13} + 2837835a_{2,15} + 64324260a_{2,17} + 1571349780a_{2,19}) \\
b_4 &= 24(a_{1,4} + 15a_{1,6} + 210a_{1,8} + 3150a_{1,10} + 51975a_{1,12} + 945945a_{1,14} + 18918900a_{1,16} + 413513100a_{1,18}) \\
&\quad (a_{2,4} + 15a_{2,6} + 210a_{2,8} + 3150a_{2,10} + 51975a_{2,12} + 945945a_{2,14} + 18918900a_{2,16} + 413513100a_{2,18}) \\
b_3 &= 6(a_{1,3} + 10a_{1,5} + 105a_{1,7} + 1260a_{1,9} + 17325a_{1,11} + 270270a_{1,13} + 4729725a_{1,15} + 91891800a_{1,17} + \\
&\quad 1964187225a_{1,19})(a_{2,3} + 10a_{2,5} + 105a_{2,7} + 1260a_{2,9} + 17325a_{2,11} + 270270a_{2,13} + 4729725a_{2,15} + \\
&\quad 91891800a_{2,17} + 1964187225a_{2,19}) \\
b_2 &= 2(a_{1,2} + 6a_{1,4} + 45a_{1,6} + 420a_{1,8} + 4725a_{1,10} + 62370a_{1,12} + 945945a_{1,14} + 16216200a_{1,16} + \\
&\quad 310134825a_{1,18})(a_{2,2} + 6a_{2,4} + 45a_{2,6} + 420a_{2,8} + 4725a_{2,10} + 62370a_{2,12} + 945945a_{2,14} + \\
&\quad 16216200a_{2,16} + 310134825a_{2,18}) \\
b_1 &= (a_{1,1} + 3a_{1,3} + 15a_{1,5} + 105a_{1,7} + 945a_{1,9} + 10395a_{1,11} + 135135a_{1,13} + 2027025a_{1,15} + \\
&\quad 34459425a_{1,17} + 654729075a_{1,19})(a_{2,1} + 3a_{2,3} + 15a_{2,5} + 105a_{2,7} + 945a_{2,9} + 10395a_{2,11} + \\
&\quad 135135a_{2,13} + 2027025a_{2,15} + 34459425a_{2,17} + 654729075a_{2,19}) \\
b_0 &= (a_{1,0} + a_{1,2} + 3a_{1,4} + 15a_{1,6} + 105a_{1,8} + 945a_{1,10} + 10395a_{1,12} + 135135a_{1,14} + 2027025a_{1,16} + \\
&\quad 34459425a_{1,18})(a_{2,0} + a_{2,2} + 3a_{2,4} + 15a_{2,6} + 105a_{2,8} + 945a_{2,10} + 10395a_{2,12} + 135135a_{2,14} + \\
&\quad 2027025a_{2,16} + 34459425a_{2,18})
\end{aligned}$$

(21)

References

- Cario and Nelson(1997). Cario MC, Nelson BL (1997) Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Department of Industrial Engineering and Management Sciences, Northwestern University
- Chen(2001). Chen HF (2001) Initialization for norta: Generation of random vectors with specified marginals and correlations. *INFORMS Journal on Computing* 13(4):312–331
- Chen and Tung(2003). Chen XY, Tung YK (2003) Investigation of polynomial normal transform. *Structural Safety* 25(4):423–445
- DeBroda et al(1989). DeBroda DJ, Dittus RS, Swain JJ, Roberts SD, Wilson JR (1989) Modeling input processes with johnson distributions. In: *Proceedings of the 21st conference on Winter simulation*, ACM, pp 308–318
- Fleishman(1978). Fleishman AI (1978) A method for simulating non-normal distributions. *Psychometrika* 43(4):521–532
- Greenwood et al(1979). Greenwood JA, Landwehr JM, Matalas NC, Wallis JR (1979) Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research* 15(5):1049–1054
- Headick(2002). Headick TC (2002) Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics and Data Analysis* 40(4):685–711
- Headick(2011). Headick TC (2011) A characterization of power method transformations through L-moments. *Journal of Probability and Statistics* 2011(2011):Article ID 497,463, 22 pages
- Headick and Sanwilowsky(1999). Headick TC, Sanwilowsky SS (1999) Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika* 64(1):25–35
- Headrick and Kowalchuk(2007). Headrick TC, Kowalchuk RK (2007) The power method transformation: its probability density function, distribution function, and its further use for fitting data. *Journal of Statistical Computation and Simulation* 77(3):229–249
- Hosking et al(1985)Hosking, Wallis, and Wood. Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27(3):251–261
- Johnson(1949). Johnson NL (1949) Systems of frequency curves generated by methods of translation. *Biometrika* 36:149–176
- Lebrun and Dutfoy(2009). Lebrun R, Dutfoy A (2009) An innovating analysis of the Nataf transformation from the copula viewpoint. *Probabilistic Engineering Mechanics* 24(3):312–320
- Liu and Der Kiureghian(1986). Liu PL, Der Kiureghian A (1986) Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics* 1(2):105–112
- Mostafa and Mahmoud(1964). Mostafa MD, Mahmoud MW (1964) On the problem of estimation for the bivariate lognormal distribution. *Biometrika* 51(3/4):522–527
- Pearson and Young(1918). Pearson BK, Young AW (1918) On the product-moments of various orders of the normal correlation surface of two variates. *Biometrika* 12:86–92
- Saxe(2002). Saxe K (2002) *Beginning functional analysis*. Springer-Verlag
- Stanfield et al(1996). Stanfield PM, Wilson JR, Mirka GA, Glasscock NF, Psihogios JP, Davis JR (1996) Multivariate input modeling with johnson distributions. In: *Proceedings of the 28th conference on Winter simulation*, IEEE Computer Society, pp 1457–1464
- Tadikamalla(1980). Tadikamalla PR (1980) On simulating non-normal distributions. *Psychometrika* 45(2):273–279
- Vale and Maurelli(1983). Vale CD, Maurelli VA (1983) Simulating multivariate nonnormal distributions. *Psychometrika* 48(3):465–471
- Zou and Xiao(2014). Zou B, Xiao Q (2014) Solving probabilistic optimal power flow problem using quasi monte carlo method and ninth-order polynomial normal transformation. *Power Systems, IEEE Transactions on* 29(1):300–306